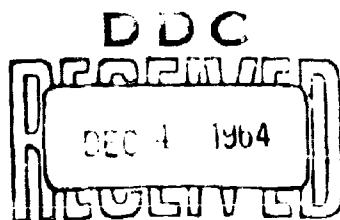SP-1735

Is Information Retrieval Now An

Established Scientific Discipline?

Ronald E. Wyllys

9 October 1964

Best
Available
Copy

$\mathbb{S}\mathbb{P}$ *a professional paper*

Is Information Retrieval Now An Established Scientific Discipline?	SYSTEM
by	DEVELOPMENT
Ronald E. Wyllys	CORPORATION
October 9, 1964	2500 COLORADO AVE.
	SANTA MONICA
	CALIFORNIA

Presented at the Nineteenth National Conference
of the Association for Computing Machinery,
Philadelphia, Pennsylvania, 25 August 1964.

ABSTRACT

This paper presents the negative case in a debate on the
question: "Resolved: Information retrieval is now an es-
tablished scientific discipline with well-defined goals,
methods, and evaluation techniques." The theme of the nega-
tive argument is that information retrieval is not a science.
Rather, it is an applied field, which deals with pressing,
practical problems. Its workers are primarily concerned with
achieving satisfactory working solutions, and only secondarily
concerned with two factors that characterize research in true
scientific disciplines--the basic understanding of some aspect
of nature and the development of theory. Some philosophical
definitions of science are quoted, and a comparison is made
between IR and what is clearly an applied field, the desalini-
zation of sea water. The paper concludes that although IR is
not yet a scientific discipline, workers in the field have at
least achieved the level at which they can recognize the basic
scientific problems underlying IR.

IS INFORMATION RETRIEVAL NOW AN
ESTABLISHED SCIENTIFIC DISCIPLINE?

My purpose is to present the negative case in a debate on the following question: "Resolved: Information retrieval is now an established scientific discipline with well-defined goals, methods, and evaluation techniques."

The crux of our debate is the definition of the words "scientific discipline." I take this phrase to be intended to imply a context of well-defined goals, methods, and evaluation techniques which have, as part of their "well-definedness," the kind of cohesiveness, coherence, and approach toward basic understanding that characterize a science. It is my contention that IR is not a scientific discipline because it is something else: namely, it is an applied field whose essence is the attempt to solve an important, pressing, practical problem by whatever means are available or can be developed. Indeed, the eclecticism of IR is one of its prime characteristics, and one which helps to distinguish it from a scientific discipline. It might seem more akin, perhaps, to an engineering discipline, but I shall later suggest an example that I think will persuade you that IR is not even a discipline within the fields of engineering.

What are the characteristics of a science or a scientific discipline? I cannot give a lecture here on the philosophy of science, but I think it is important to consider a few things about what a science _is_, about what a science is _not_, and about what _is not_ a science.

Braithwaite [1] says: "The function of a science...is to establish general laws covering the behavior of the empirical events or objects with which the science in question is concerned, and thereby to enable us to connect together our knowledge of the separately known events, and to make reliable predictions of events as yet unknown."

Ackoff [2], even more pertinently to the point at issue in our debate, says: "We can consider science as a process of inquiry; that is, as a procedure for (a) answering questions, (b) solving problems, and (c) developing more effective procedures for answering questions and solving problems." By contrasting question-answering and problem-solving, Ackoff is leading up to the point of my argument. He goes on to say: "Not all inquiry is scientific. There is a large class of nonscientific inquiry, including what we call 'common sense' inquiry....It has been proposed in the past that common-sense...inquiry is concerned with more immediate and practical problems than is science. It has become increasingly clear, however, that science, particularly 'applied science,' does deal with immediate and pressing problems."

The distinction, then, between science--in particular, applied science--and the common-sense kind of problem-solving to which I maintain IR belongs, lies not so much in the practicality of the problems as in (so I am arguing) the highly eclectic nature of the solutions achieved or attempted. But this is

not the only distinction. I should like to try to bring out at least one more kind of distinction by referring once again to a philosopher of science.

Mehlberg [3] characterizes a science as a collection of data, theories, and validation methods. He defines a theory as a coherent and cohesive set of laws. Laws must be pertinent and unambiguous answers to the question under consideration. The laws must be supported by the available data, which must be sufficiently reliable, numerous, and varied to furnish adequate support. A theory should consist, says Mehlberg, of "relatively few explicitly formulated statements from which a theoretically infinite set of propositions can be derived by a specifiable logical procedure."

But according to Mehlberg a theory must consist of answers to questions, and these answers must be supported by data. If a theory is to contain a theoretically infinite set of propositions, how can they all be supported by the finite set of data offered in the presentation of the theory? The way out of this transfinite difficulty, according to Mehlberg, is that the scientist offers, along with his theory and data, a description of the methods he has used in arriving at his answers--methods which can then serve as validation procedures for testing any desired one of the theoretically infinite set of propositions. that are logically derivable from the theory. Mehlberg holds that these validation methods "form the gist of the whole method of science" (a strong statement indeed, which he devotes most of his book to justifying). If a validation method turns out also to have predictive power, that is, to answer questions besides those it was designed to answer, then so much the better.

Thus Braithwaite, Ackoff, and Mehlberg agree that a science must have at least the following characteristics:

a) it must provide answers to questions;

b) the answers must be consistent with empirical data;

c) there must be a unifying structure that incorporates the answers and makes it possible to extend them; and

d) predictions should be possible, and if they exist, must be testable.

I have cited these philosophers of science to add the weight of authority--for whatever that may be worth--to my claim that IR is not a science but instead is an area of application of eclectic common sense: it is, at worst, "a bag of tricks," but better--and I think more accurately--it is a set of useful techniques for achieving a practical goal. I think IR research can be viewed as having both an intermediate subgoal and a final goal. The intermediate subgoal is that of improving methods of document retrieval; the final goal is to achieve what is called--with a nebulousness that is, at least for the present, unavoidable--"fact retrieval."

It is certainly true that achieving these two goals will involve progress in
some indisputably scientific disciplines. Chief among these, I think, will be
progress in achieving an adequate understanding of the process of verbal com-
munication. Progress here will surely involve the disciplines of linguistics
and mathematics, and may involve other disciplines before the job is completed.
One might also expect that progress in the "hardware" areas of science will
sooner or later be applicable to some of the equipment problems of IR; for ex-
ample, cryogenics may come to have an application here. But I think it is also
certainly true that IR-related work in such fields has so far been carried out
mostly by persons who do not consider themselves to be information retrievalists.
Rather, most people in IR have come from either the computer field or from the
library field. They have come into IR in order to solve immediate, pressing,
practical problems. Their primary motivation is to achieve certain practical
goals rather than to develop theories of IR or to create a logical entity having
the characteristics of question-answering, consistency with empirical data, a
unifying structure, and testable predictions--the characteristics that Braithwaite,
Ackoff, and Mehlberg agree on in defining a science. The distinction between a
scientific discipline and a problem-solving effort is illustrated here by the
distinction between the attempts to gain basic understanding of how verbal com-
munication is accomplished, and the attempts to build this or that document re-
trieval, sentence retrieval, or formatted-fact retrieval system.

At the risk of being tedious, let me repeat that the distinction lies in whether
the goal is primarily a pragmatic one--to achieve a functioning, man-machine
system that satisfies some practical, pressing need--or whether the goal is to
achieve an ever deepening understanding of some facts of nature by creating or
extending a structure of intimately interrelated fact and theory. Only the
latter goal can justify the label of "scientific discipline."

As a further illustration of the distinction, let us consider another area,
which I am sure no one would call a scientific discipline, and which I think
is not even an engineering discipline in the sense that, say, hydraulic engi-
neering is a discipline. I have in mind the problem of the desalinization of
sea water.

The desalinization problem shares several characteristics with IR. It is a
pressing problem; it needs practical solutions; the cost of possible solutions
is a very important factor in judging them; and techniques of many kinds, from
many fields, are being tried out as possible solutions. We can get an idea of
the variety of techniques by excerpting from the table of contents of a recent
report [4] from the Department of the Interior on techniques under study in
1962 as possible means of desalinization:

> Distillation Processes
> > Multistage flash
> > Vapor reheat flash and liquid-liquid heat exchange
> > Thin film vapor compression
> > Wiped film--fluted tube thin film

       Fluidized bed
       Absorption--multistage flash
       Heat pump

Humidification Processes
   Multiple-effect humidification
   Solar stills
   Diffusion still

Membrane Processes
   Electrodialysis
   Reverse osmosis
   Vapor gap osmosis

Ion Exchange Processes

Freezing Processes

Hydrate Processes

Solvent Extraction Process

Other applied research [including algae, high-frequency sound, and high-frequency electric currents]

Compare, if you will, this long list of techniques that hopefully may prove useful in desalinization with the many and varied techniques that have been used or tried in IR. (I shall refrain here from listing any IR techniques, for two reasons: first, anyone in the field is well aware that the techniques have been many and varied; and second, it would be unfair, to persons who have associated themselves with any of the techniques I might give as examples, to single out any techniques for mention in the present context.) Such a comparison will confirm, I think, the eclectically pragmatic orientation of most current work in IR.

All this is not to say that IR cannot become a scientific discipline. On the contrary, I think it can and will. There is, after all, a truly scientific problem underlying IR. Indeed, it is the enormous difficulty of this problem that, combined with the great pressures for at least symptomatic solutions, has thus far kept IR from becoming a scientific discipline. This basic problem is that of the nature of information, especially information conveyed in the form of natural language. It is one thing to transmit information from a sensor to a reactor, as, for example, when a thermocouple senses that the water in a heater has reached a temperature of 180° F. and sends out a signal that shuts off the fuel supply. It is another thing entirely when I utter the sentence, "What is the maximum safe working temperature of this water heater?" We do not yet have the tools to understand the complex of difficulties that such a sentence

obtrudes against attempts to specify a procedure for analyzing it rigorously.
But we have at least come to recognize what such tools will require: viz.,
an understanding of the intricate relationships among the information user,
the written or spoken word, and the real world. As we progress toward achieving
that understanding, IR can become a scientific discipline.

REFERENCES

[1] Braithwaite, R. B. Scientific Explanation. Cambridge: Cambridge University Press, 1953, p. 1.

[2] Ackoff, R. L., Gupta, S. K., and Minas, J. S. Scientific Method. New York: Wiley, 1962, p. 1-2.

[3] Mehlberg, H. The Reach of Science. Toronto: University of Toronto Press, 1958, Chapters 1 and 2.

[4] U. S. Department of the Interior. Saline Water Conversion Report for 1962. Washington: Government Printing Office, 1963, p. viii-ix.

System Development Corporation,
Santa Monica, California
IS INFORMATION RETRIEVAL NOW AN
ESTABLISHED SCIENTIFIC DISCIPLINE?
Scientific rept., SP-1735, by
R. E. Wyllys. 9 October 1964, 6p.

Unclassified report

DESCRIPTORS: Information Retrieval.

Presents the negative case in a debate
on the question: "Resolved: Information
Retrieval is now an established scientific
discipline with well-defined goals, methods,
and evaluation techniques." Sets forth

---

that information retrieval is not a
science but an applied field which
deals with pressing practical problems.
Reports that its workers are primarily
concerned with achieving satisfactory
working solutions, and only secondarily
with the basic understanding of some
aspect of nature and the development of
theory. Presents comparisons between IR
and other applied fields. Concludes
that although IR is not yet a scientific
discipline, workers in the field have
achieved the level at which they can
recognize the basic scientific problems
underlying IR.